"Express Mail" Mailing Label No. **EV 352470144 US**

PATENT APPLICATION
ATTORNEY DOCKET NO. OR03-10201

METHOD AND APPARATUS FOR PROTECTING PRIVATE INFORMATION WITHIN A DATABASE

Inventor: Vipin Samar

BACKGROUND

Field of the Invention

[0001] The present invention relates to databases. More specifically, the present invention relates to a method and an apparatus for protecting private information within a database.

Related Art

[0002] Many organizations collect personal private information from individuals for various reasons, and store the information in a database. If this information should fall into the wrong hands, the information could be used to the detriment of the individuals.

[0003] For example, many organizations use an individual's Social Security Number (SSN) as an identifier for an individual, and also as a primary key in their database. This can be a problem because SSNs are one of the primary pieces of information used for identity theft. These SSNs, along with other personal private information, aggregated in a database make a compelling target for hackers and thieves. Consequently, many localities have passed laws forbidding organizations from using SSNs or not allowing them to store SSNs in plain text.

[0004] In many cases the SSN is not even useful as information to the organizations. The organizations simply use the SSN as a unique key to look up information associated with an individual in a database. SSNs are convenient to use for this purpose because they are guaranteed to be unique and most individuals have their SSN committed to memory.

[0005] One way to protect private information is to encrypt the private information before it is stored in a database. In this way, even if someone illegally accesses the encrypted data, they will not be able to use it. However, if an application is able to access encrypted information within a database, the keys for encrypting and decrypting the data must be located somewhere on or near the server. Otherwise, the encrypted data would be useless to the application. Because the keys are located so that they can be accessed directly or indirectly by the application, the database administrator and possibly the programmers also have access to the keys, and consequently, have access to the encrypted data.

[0006] Hence, what is needed is a method and an apparatus for securing private information in a database without the problems listed above.

SUMMARY

[0007] One embodiment of the present invention provides a system that facilitates protecting an item of private information in a database, wherein the item of private information is used as a key for retrieving data from the database. During operation, the system receives the item of private information and creates a hash of the item. The system then stores the hash in the database along with any associated information in a database record containing the hash.

[0008] In a variation on this embodiment, creating the hash involves creating a SHA-1 or MD5 hash.

[0009] In a variation on this embodiment, the hash is created automatically by the database in a manner that is transparent to an application, which manipulates the private information.

[0010] In a variation on this embodiment, processing a query involving the item of private information involves creating a hash of the item of private information, and querying the database using the hash.

[0011] In a variation on this embodiment, the item of private information can include one of: a social security number, a driver's license number, a passport number, an email address, a person's name, and a person's mother's maiden name.

[0012] In a variation on this embodiment, multiple items of private information can be combined prior to creating the hash.

[0013] In a variation on this embodiment, creating the hash further involves checking a column attribute in the database to see if "privacy" is enabled, and if so creating the hash.

[0014] In a variation on this embodiment, the database is a Lightweight Directory Access Protocol (LDAP) database.

BRIEF DESCRIPTION OF THE FIGURES

[0015] FIG. 1 illustrates exemplary databases in accordance with an embodiment of the present invention.

[0016] FIG. 2 presents a flowchart illustrating the process of protecting private information in accordance with an embodiment of the present invention.

[0017] FIG. 3 presents a flowchart illustrating the processing of a query in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

[0018] The following description is presented to enable any person skilled in the art to make and use the invention, and is provided in the context of a particular application and its requirements. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art, and the general principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present invention. Thus, the present invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.

[0019] The data structures and code described in this detailed description are typically stored on a computer readable storage medium, which may be any device or medium that can store code and/or data for use by a computer system. This includes, but is not limited to, magnetic and optical storage devices such as disk drives, magnetic tape, CDs (compact discs) and DVDs (digital versatile discs or digital video discs), and computer instruction signals embodied in a transmission medium (with or without a carrier wave upon which the signals are modulated). For example, the transmission medium may include a communications network, such as the Internet.

Exemplary Databases

[0020] FIG. 1 illustrates exemplary databases 100 and 110 in accordance with an embodiment of the present invention. Database 100 and database 110 contain the same information. However, database 100 is an unencrypted database comprising table 102, whereas database 110 is a privacy-enabled database comprising table 112. As illustrated in FIG. 1, Social Security Numbers (SSNs) in table 102 are stored in plain text. Hence, anyone who has access to table 102 can view all of the information in table 102, including the SSNs. In contrast, table 112 contains the same data as table 102, but with hashed SSNs. Note that generating the hash can involve performing any one of a number of one-way functions, including SHA-1 and MD5.

Process of Protecting Private Information

[0021] FIG. 2 presents a flowchart illustrating the process of protecting private information in accordance with an embodiment of the present invention. The process starts when the system receives a piece of private information to be stored in the database (step 202). This piece of private information is typically a Social Security Number, Driver's License Number, or some other unique piece of information that is used as a key in the database to look up a record associated with an individual. Next, the system checks a type value for a column in which the information is to be stored in database 110 (step 204), and if this type value indicates that privacy is enabled for the column, the system creates a hash of the private information (step 206). At this point, the system also throws away the private information, or alternatively stores it in a secure location (step 207). As mentioned previously, any type of one-way function can be used to generate the hash. Also note that several pieces of private information can be combined into

the hash. Once the hash is created, the hash and other related information is stored in a record in database 110 (step 206).

[0022] In one embodiment of the present invention, the hash is created automatically by the database in a manner that is transparent to the application. A new column attribute can be defined in the database instructing the database to always hash values upon inserting the values into the column. Note that performing this hashing automatically provides security without having to modify applications that access the database. These hash values can also be indexed to speed lookups. However, range searches become complicated. One possible method for performing a range search is to generate each value in the range, perform a hash on each value, and then look up each hash in the database.

Processing a Query

[0023] FIG. 3 presents a flowchart illustrating the processing of a query in accordance with an embodiment of the present invention. The system starts when the system receives a query that involves the piece of private information (step 302). Next, the system checks the associated column type in database 110 (step 304). If this column type indicates that privacy is enabled, the system creates a hash of the piece of private information (step 306). The system then performs the query using the hash in place of the piece of private information (step 308). For example, the system can perform a "select" on the database where the hash is substituted in the "where" clause in place of the piece of private information. Note that as described previously, the hashing can take place at the database level in a manner that is transparent to the application. For example, the select statement might contain the private information in the "where" clause, and the database, knowing that the column referenced by the "where" clause is marked

as privacy-enabled, would automatically hash the data before performing the lookup.

[0024] In one embodiment of the present invention, the hashing operations are not performed automatically, but are instead performed by a programmer. In this embodiment, the methods for checking if a column is privacy enabled and for creating the hashes are exposed to programmers through an API.

[0025] The foregoing descriptions of embodiments of the present invention have been presented for purposes of illustration and description only. They are not intended to be exhaustive or to limit the present invention to the forms disclosed. Accordingly, many modifications and variations will be apparent to practitioners skilled in the art.

[0026] For example, although the present invention is described in the context of a relational database, the present invention is not limited to relational databases. In general, the present invention can be applied to any type of database, including relational databases, hierarchical databases, centralized databases and distributed databases. In one embodiment of the present invention, the present invention is used to hash directory information stored in a Lightweight Directory Access Protocol (LDAP) database.

[0027] Additionally, the above disclosure is not intended to limit the present invention. The scope of the present invention is defined by the appended claims.